Single Numeric Evaluation Matrix

* Using Dev set + Single real number evaluation matrix can speedup iterating.

precision and recall:

explained with examples:

precision: of all predicted cats, what % actually are cats. (TP/TP+FP)

recall: of all actual cats, what % are recognised. (TP/TP+FN)

F1 score: "Average (Harmonic mean) of precision and recall.

$$= \boxed{\frac{2}{\frac{1}{P}+\frac{1}{R}}}$$

This is a good single numeric evaluation matrix.

satisfying and optimizing metrix

It's always nice to have a single numeric number as your goal to optimize your model, but sometimes we have to cousider other critirias at the same time, in such cases, we need to set

Satisfying matrix

N matrix : 1 optimizing — make sure that there's still only 1 numeric optimizing matrix

N-1 satisfying

e.g. maximize accuracy — optimizing matrix

subject to running time ≤ 100 ms — satisfying matrix

train, dev, and test set.

Guide line: Choose a dev set and test set to reflect data you expect to get from future. (from same distribution)

Size of dev set: set dev set to be big enough to detect differences in algorithm/models you are trying out.

Size of test set: set your test set to be big enough to give high confidence in the overall performance of your system.

* if doing well on your metrix + dev/test set does not correspond to doing well on your application, change your metric and/or dev/test set.